Visual analysis design to support research into movement and use of space in Tallinn: A case study

Qiuju Zhang[1], Aidan Slingsby[2], Jason Dykes[2], Jo Wood[2], Menno-Jan Kraak[1], Connie A. Blok[1] and Rein Ahas[3]

Abstract

We designed and applied interactive visualisation to help an urban study group investigate how suburban residents in the Tallinn Metropolitan Area (Estonia) use space in the city. We used mobile phone positioning data collected from suburban residents together with their socio-economic characteristics. Land-use data provided geo-context that helped characterise visited locations by suburban residents. Our interactive visualisation design was informed by a set of research questions framed as identification, localisation and comparison tasks. The resulting prototype offers five linked and coordinated views of spatial, temporal, socio-economic characteristics and land-use aspects of data. Brushing, sorting and filtering provide visual means to identify similarities between individuals and facilitate the identification, localisation and comparison of patterns of use of urban space. The urban study group was able to use the prototype to explore their data and address their research questions in a more flexible way than previously possible. Initial feedback was positive. The prototype was found to support the research and facilitate the discovery of patterns and relations among groups of participants and their movements.

Keywords

Interactive visual analysis, multiple coordinated views, user questions, user tasks, movement, geo-context

Introduction

Over the last decade, rapid suburbanisation in Tallinn (the capital of Estonia) has changed the socio-economic composition of the population in different parts of the metropolitan area.[1] Among other things, this has led to a marked increase in commuting behaviour between the suburbs and the city centre,[2] having a profound effect on the relationship between residents and their use of urban space,[3] with obvious implications for planning and service provision. This rapid development was largely ad hoc and unregulated, and thus not supported by infrastructural improvements. Geographers and urban planners investigating these changes lack the official sources of data to do this – the last census was in 2000 – and are turning to alternative sources of data, with individual-level surveys and movement data among those being investigated.[4]

The movement of individuals is influenced by characteristics of the local surroundings,[5] termed 'geo-context'. When analysing these data, incorporating geo-contextual data can help researchers understand

[1]University of Twente, Enschede, Netherlands
[2]City University London, London, UK
[3]University of Tartu, Tartu, Estonia

Corresponding author:
Qiuju Zhang, University of Twente, Hengelosestraat 99, Enschede, 7500 AE, Netherlands.
Email: zhang16560@itc.nl

the motivations of movement. For example, the apparent effects of tourist attractions on tourists' behaviours can help planning, making changes to improve tourist experiences or to generate more revenue. Interactive visualisation that takes advantage of powerful human cognitive capabilities is well suited to support movement analysis that takes geo-context into account.[5–7]

Having a deep knowledge of the aims of researchers and their research questions are important prerequisites for designing appropriate interactive visualisation to support their research.[8] We collaborated with an urban study group, studying how new suburban residents connect to and use space in the city. The urban study group collected socio-economic data and mobile-phone-derived locational data at fixed temporal sampling intervals from 277 of these residents. Using land use as the geo-context, research questions focused on temporal aspects of the connections between the suburbs and the city centre and land-use consumption and how it relates to socio-economic characteristics of suburban residents.[9]

Our aim was to work closely with the urban study group to understand their research questions, design appropriate interactive visualisation to assist them in answering their research questions and to implement the visualisation. The novelty of this work is in the way in which we, in close collaboration with users, translated their research requirements into interactive visualisation design. We describe the process we used to achieve this, reflect on how successful it was and consider whether these data and these visualisation techniques can be applied more widely to urban planning research.

Related work

Previous studies that used the most recent census (2000) and socio-economic information about suburban residents have confirmed that during the 1990s, there was an increase in commuting between suburban communities and Tallinn city centre.[10] Kährik and Tammaru[1] analysed changes of suburban population using survey data from the Household Panel Survey (2004) and the New Residential Areas Survey (2006). Their results show that young, wealthy and well-educated households moved from the city to new suburban settlement areas after 2000. Studying these new suburban residents' spatial and temporal dynamics helps us understand the impact of these changes.

The widespread popularity and usage have made mobile phones suitable vehicles to collect information on individuals' movements, passively (stored on the device to be retrieved later) or actively (retrieved regularly over the network with the user's permission).[11] Researchers from the University of Tartu used active mobile positioning data collected from new suburban residents to study their temporal connections to Tallinn city centre through analysing hourly, daily and weekly changes in their travel distances.[9] They inferred home and work locations from these but did not consider geo-context.

Visual analysis of mobile phone data has been conducted for mining human movement patterns, under the time geography framework[12] and using a visual analytics framework.[13] Several researches used visual analysis of mobile phone activities to construct social networks.[14,15] As different land uses have different degrees of human clustering,[16] consideration of land use can help explain human behaviour.[17] Research has confirmed that individual activity patterns are strongly related to land-use patterns.[18] However, in these studies, there is neither explicit collaboration with domain users nor designs to support analysis guided by users' questions or tasks.[8]

In the visualisation domain, researchers have categorised generic task typology from different perspectives, helping us design techniques that fit the task at hand. MacEachren[19] summarised seven types of questions for querying spatio-temporal data. Knapp took a cognitive perspective, reducing the 11 tasks suggested by Wehrend and Lewis[20] into 4: identify, locate, compare and associate.[21] Andrienko and Andrienko used level of reading[22] to categorise elementary and synoptic tasks.[23] Friendly and Kwan[24] demonstrated the importance of display configuration being related to the task at hand. Slingsby et al.[25] explored the effects of selecting layouts for addressing research questions. Although general task descriptions are available, the task-based design of views and the matching with application-driven questions have rarely been investigated by the current researches.[26]

Scenario, data and task analysis

Scenario

The suburbanisation during the real estate boom in Tallinn was rapid because of the need for better housing and easier and cheaper opportunities to develop residential areas outside the city.[10] The growing wealth of households and lower mortgage interest rates over the period also contributed to the rapidity of development of new residential areas.[1] However, transport and public services did not keep pace with these changes as the authorities had a lack of financial instruments and did not have the legal framework to regulate the boom.

Studies have shown that young (20–35) and educated (53% with university degree) families moved to the new suburbs from Tallinn city.[1] Official sources of

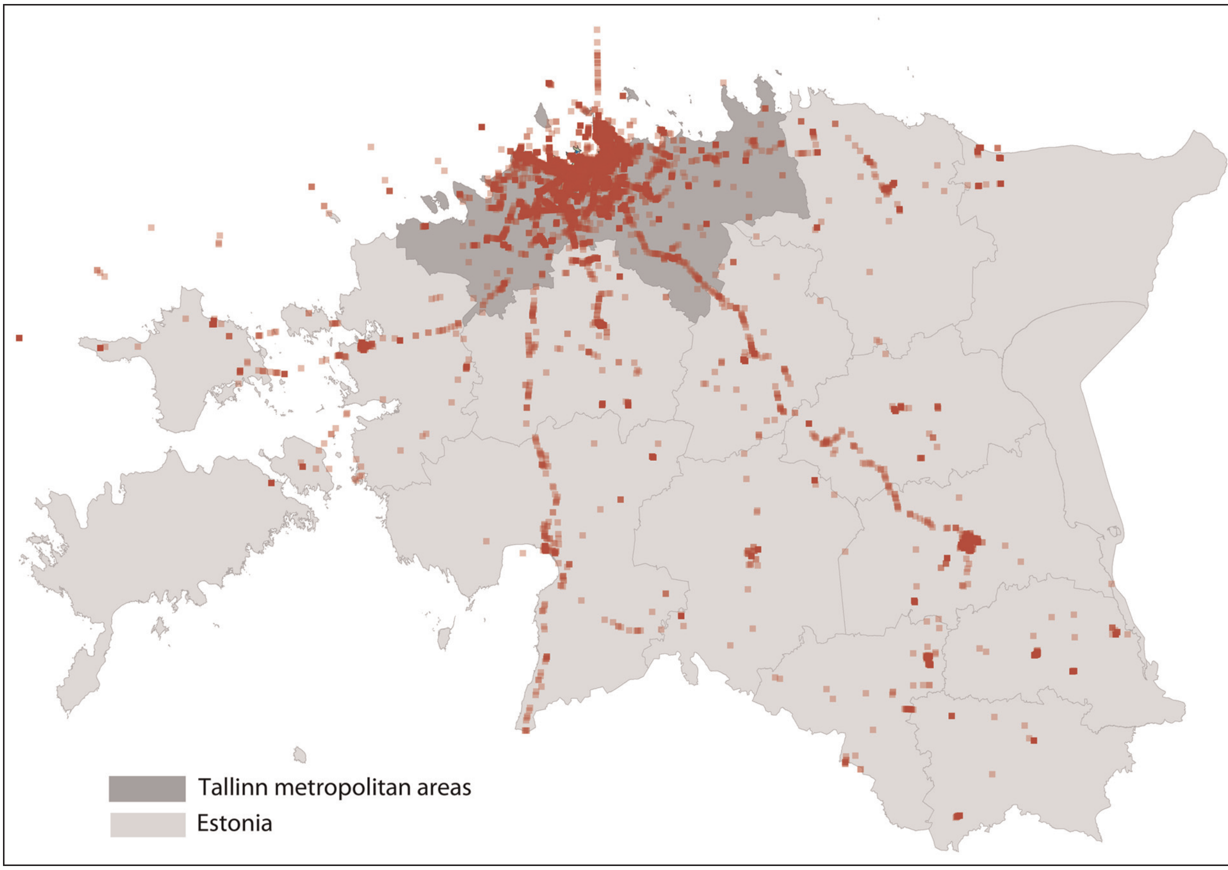

Figure 1. Over 140,000 mobile phone positioning locations of 277 residents in the Tallinn Metropolitan Area over a period of 8 days. The level of opacity indicates the density of points.

data are inadequate for studying this redistribution of population, because government statistics and public registers have failed to handle such new settlements in Estonia and the last census was before the boom. There are socio-economic and transport-related implications that the urban study group wish to understand better.

Data

Active positioning data from mobile phones were gathered over 8 days (5–12 April 2006) from 277 participants living in the Tallinn Metropolitan Area (Figure 1). The data have spatial coordinates, time stamps and socio-economic characteristics of phone users, including gender, employment status, age, income and education, summarised in Table 1. The data can be generalised into two measurement levels: nominal and higher than nominal, and the last category for convenience is referred to as 'ordinal' in the following sections. The data have a spatial accuracy of half a kilometre and are sampled at 15-min intervals between 06:00 and midnight and at 2-h intervals outside this period. In total, over 140,000 locations were collected.

We were supplied with a land-use classification scheme of 15 categories[27] in Tallinn city centre (Figure 2). We combined 'defence and military', 'harbour' and 'special functions' and assigned each to a unique colour using a ColorBrewer qualitative colour scheme.[28]

We identified four components of the data: participants as moving objects (o), space (s), time (t) and land-use geo-context (l). These components form patterns (p; Figure 3).

User questions and task analysis

We took a user-centred approach to the design by working closely with the urban study group.[29,30] Through a number of discussions, we established the context of their research and the types of research questions they had. We then designed two-dimensional (2D) and three-dimensional (3D) graphical examples that depicted different aspects and characteristics of their data. We used these examples in a subsequent

Table 1. Data attributes.

Components	Attributes	Categories
Participants	Gender	Male, female
	Age	10–67
	Income	<4000, 4001–8000, 8001–12,000, 12,001–16,000, 16,001–20,000, $>20{,}001$
	Employment status	Pupil, student, wage worker, businessman, homemaker, unemployed, retired
	Education	Secondary education or lower Vocational education Higher education
Space	Coordinates	*x,y*
	City/suburbs	City centre (defined by the administrative borders of Tallinn, covering the same area of the land-use map in Figure 2), suburban area
Land use	Land use	Land-use units and types according to Estonian Land Board (Figure 2)
Time	Time moment	

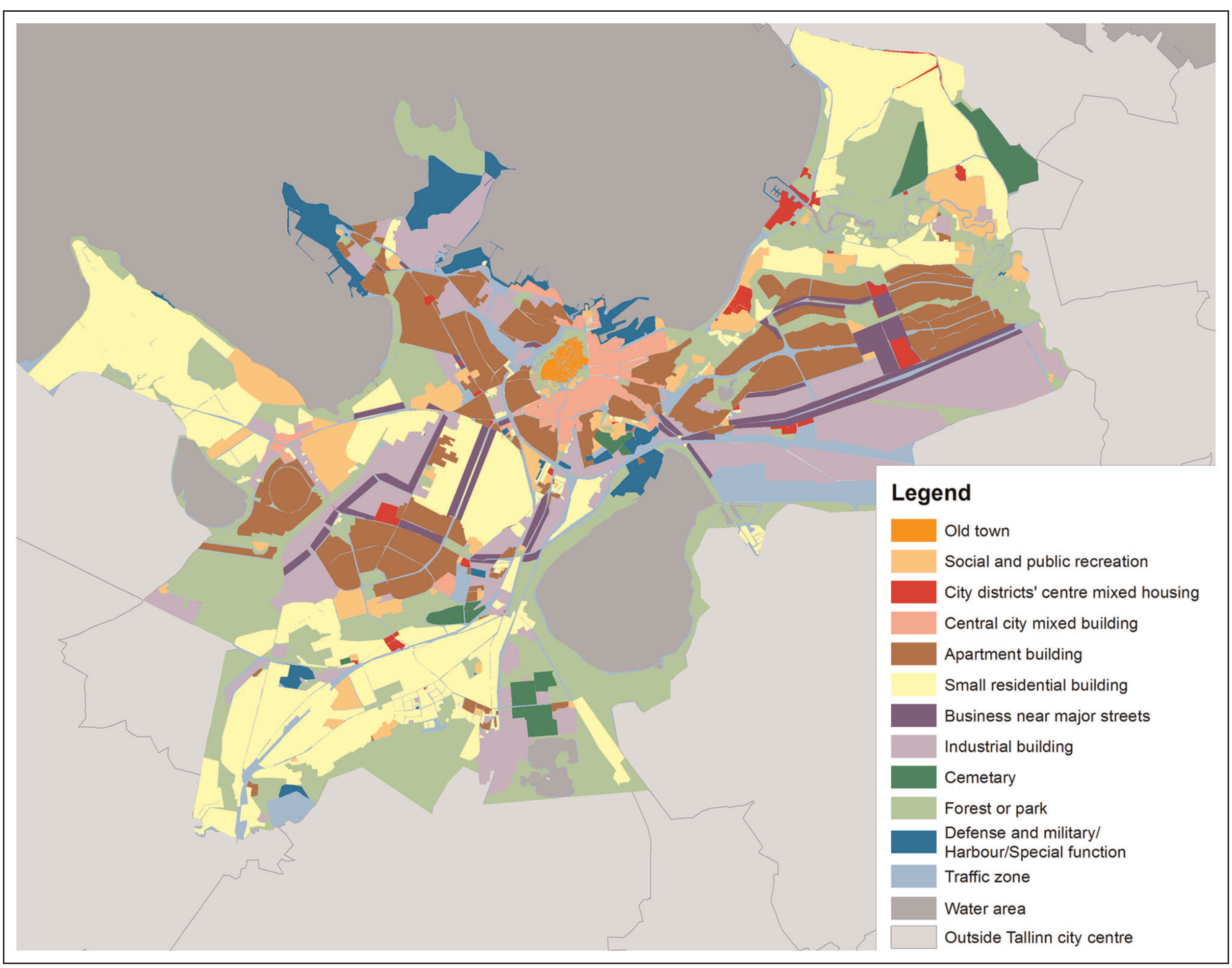

Figure 2. Land-use classes in Tallinn city centre.[27]

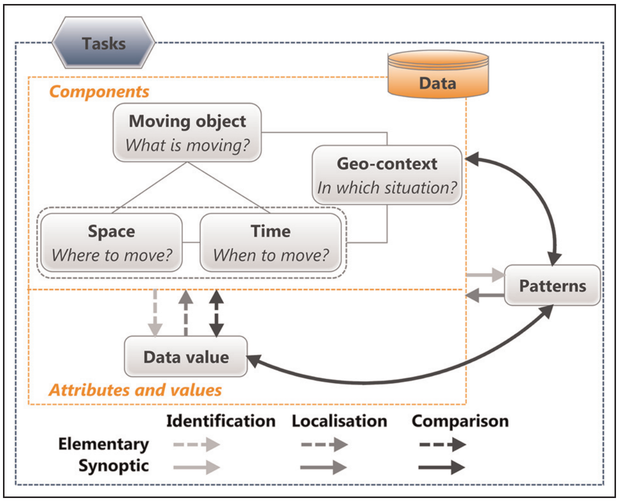

Figure 3. The three task types of identification, localisation and comparison at both elementary and synoptic levels are classified according to relationships between data components, values and patterns. Arrows indicate the starting point and the target.

workshop to demonstrate possibilities, facilitate discussions and help stimulate the formulation of more research questions.[29,30] The outcome of the workshop was a set of questions that the urban study group found difficult to answer with their existing toolset. These questions are listed later in this section.

Based on an existing task taxonomy[21,23] and an analysis of the relationships between data and pattern, we categorised the questions into three primary user tasks: identification, localisation and comparison, each of which can exist at an elementary or synoptic level (Figure 3). Elementary tasks address individual data attributes and values, whereas synoptic tasks look for trends and overviews.[23] A task usually implies two parts: a starting point and a target to be sought or compared with.

Identification tasks search for specific data values (elementary, also called 'direct lookup'[23]) or patterns that describe trends or associations in a subset of data (synoptic, also referred to as 'pattern identification'[23] or 'association'[21]). *Localisation tasks* look for when and where objects (elementary) or patterns (synoptic) exhibit particular characteristics (also referred to as 'inverse lookup' or 'pattern search'[23]). *Comparison* determines which relations (e.g. similarities or differences) exist between data attributes/values (elementary) or patterns (synoptic). Some researchers identify the additional 'relation-seeking' task type that looks for relations between data components, but we consider this to be a composite of the other task types.[31]

We classified the research questions identified at the workshop according to task type and list these below. Starting points and targets for questions are identified with bold and italics, respectively.

Q1. How many persons have regular connections to the city centre?
Synoptic task: Identification of **participants (o)** who have *temporal periodicity (p)* of **connection to the city centre (s,t)**.
Q2. How often do suburbanites travel to city centre?
Synoptic task: Identification of **participants' (o)** *temporal frequencies (p)* of **connection to the city centre (s,t)**.
Q3. Which connections to the city centre do persons have during leisure time (after regular business hours and during weekends)?
Elementary task: Localisation of **participants' (o)** *visits to the city centre (s/l)* **during leisure time and weekends (t)**.
Q4. Which geographical locations and land-use types are associated with the jobs and schools of suburbanites?
Synoptic task: Localisation of *where (s) and in which land uses (l)* are **participants (o) (co)-located (p) during office hours (t)**.
Q5. How complex (diversity of routes and destinations) are regular trips to the city centre?
Elementary task: Comparison of *destinations (s) and land uses (l)* of **participants' trips**.
Q6. How does the connection of persons to the city centre vary over time?
Synoptic task: Comparison of *differences/changes (p)* of *participants' (o) connections to the city centre (s)* **over time (t)**.
Q7. Can we infer different occupations by the nature of their visits to the city centre?
Synoptic task: Identification of *visiting patterns (p)* of **different groups of participants' (o) movements (s/l, t)**, then comparison of *these (p)*.

Design

Information gathered from the workshop at which we presented our examples and the research questions that resulted were used to inform our design.

The diversity of research questions suggests that the visualisation design needs to be able to *depict multiple aspects* of the data. However, we established at the workshop that it is important for users of the visualisation to *understand exactly which information is being presented* to them.[32] We therefore decided to use multiple coordinated views[33] to depict multiple aspects of the data, where each individual view remains in the same area of the screen and uses visual variables consistently. The research questions suggest that relationships

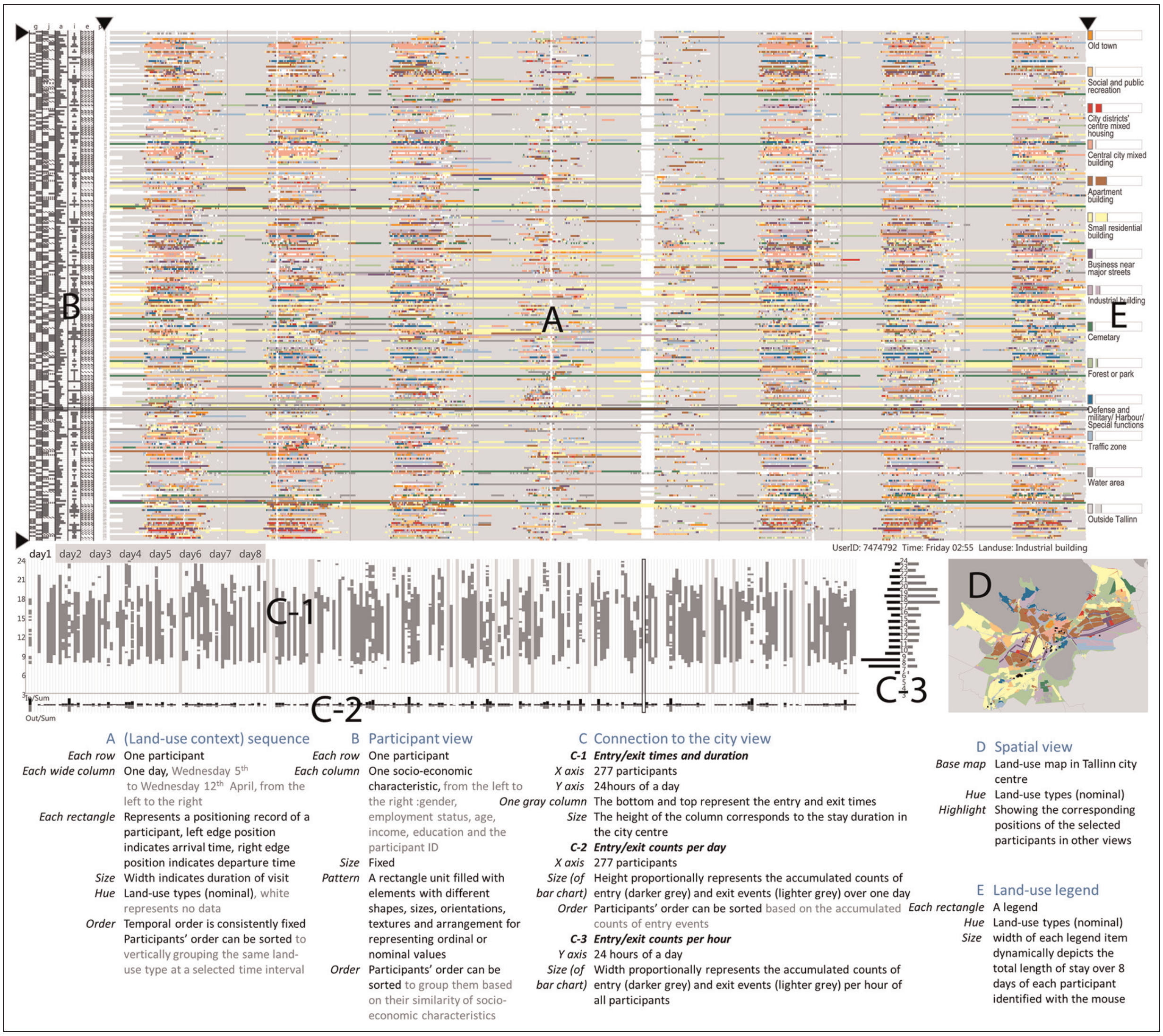

Figure 4. Screenshot of the prototype that implements our design as described in the 'Design' section.

between different subsets of the data are important. Our experience and that of others suggest that trends and comparisons are easier when targets of the comparison are visible concurrently.[34] For these reasons, our overall design uses five linked and coordinated views of spatial, temporal, social characteristics and land-use aspects of the data (Figure 4), showing all these aspects on a single screen. This design was guided by the data characterisation and user tasks informed by the visual encoding taxonomy described below.

Visual encoding and interaction

Key to good visualisation design is the appropriate visual encoding of data for the data characteristics and the task at hand. We used the advice[19,22,35] summarised in Table 2 to help us map visual variables to aspects of the data based on the characteristics of the data. Task type also influences decisions on visual encoding. Bertin[22] suggested that visual variables can be independently judged on 'selectivity' and 'associativity' (Table 2). A selective variable enables map readers to quickly isolate all the correspondences belonging to the same category.[22] An associative visual variable provides same visual 'weight',[19] supporting equalisation and grouping of all the categories during perception.[22] Some visual variables are both associative and selective if applied to ordinal/nominal data (shading in Table 2).

Bertin recommended that an elementary task should use a selective variable, whereas a synoptic task should use an associative variable. However,

Table 2. Effectiveness of visual variables related to *perception of associativity and selectivity* (two leftmost columns) and *measurement level* (two rightmost columns).

Visual variables	Associative	Selective	Nominal	Ordinal
Locations (in the plane)	•	•	•	•
Size		•		•
Colour value		•		•
Colour saturation		•		•
Colour hue	•	•	•	
Texture	•	•	•	•
Orientation	•	•	•	
Shape	•		•	
Arrangement	•	•	•	

"•" = effective
Source: Derived from Bertin,[22] Morrison[35] and MacEachren.[19]
Variables that are both associative and selective for ordinal/nominal data are shaded.

identification and localisation tasks in both elementary and synoptic levels require targets to become salient[36] so that users can find targets quickly, making selective variables more suitable. Comparison tasks are best supported when data are depicted in a perceptually uniform manner[31,37] to avoid visual bias, making associative variables more suitable. Three types of tasks at elementary and synoptic levels have interwoven relationships with the selection of selective or associative visual variables. For a view intended to perform all tasks, the primary choices of visual variables are those that are both associative and selective. This helps ensure that users perceive attributes or values that are sufficiently similar to be treated as one unit and those that are sufficiently different to be visually separated.

Where multiple views exist, well-designed interaction helps relate data across views, for example, through brushing,[38] sorting[24] and filtering. Interaction[39] can be added to support Shneiderman's[40] 'overview, zoom & filter and details-on-demand' visual exploration activities. The design of each view with interaction will now be discussed in turn.

(Land-use context) sequence view

In Figure 4(A), we are treating time as a one-dimensional (1D) sequence. We map this to one of the planar screen dimensions (x-axis), freeing the other axis to distinguish between participants. The result is a sequence view that shows the sequence of land-use types visited in chronological order (Figure 4(A)). Land-use visits are represented as rectangles whose colour indicates land-use type (nominal), left edge position indicates arrival time, right edge position indicates departure time, width indicates duration of visit and row number indicates the participant's identity (Figure 5(B)). As a visualisation technique, it shares characteristics of treemaps,[41] trellis plots[42] and mosaic plots[43] and helps perform multiple tasks and provides answers to questions Q3–Q5 and Q7. The view produces consistent small-multiple-like arrangements that enable repeated patterns to be visually identified and compared,[44] important for Q7.

Although participants have no inherent sort order (as in Figure 4), they can be sorted by any aspect of interest (such as income) in accordance with a research question. Sorting rows places the land-use sequences for similar participants closer together, helping visually determine whether sequences vary by participant type (Q7). Simply cycling through different sorting orders can help quickly scan for patterns.[24] Although ordinal values have an inherent order, nominal values do not. We do try and assign a meaning order to nominal categories where possible – often using an alphabetical order that helps facilitate lookup – but as there is no inherent order, sorting only has the effect of grouping participants.

We use *multiple sorting* in which items are sorted by different aspects in turn. During sorting, when participants share the same category, they are juxtaposed in the same order as in the previous sort.

Participants can be sorted on socio-economic characteristics as well as on the land-use type visited at the time indicated with the mouse cursor. For example, in Figure 8, participants are sorted by their land-use type visited at 10:00, then 19:00 and then 02:00 on 1 day. These times were significant for one of the research questions and it places participants who occupied the same type of land use at these three different times together. Sorting by connections to the city centre is also possible.

Other interactive techniques are available. Users can use brushing[38] to identify participants in other views. Details-on-demand for each rectangle are available at the bottom of the sequence view. Participants can be filtered by time (by dragging the two triangles horizontally along the top of the screen) or by row (by dragging the leftmost triangles vertically) after an appropriate sorting (Q3 and Q4).

Although the sequence view does not encode geographical information, we provide interaction that identifies land-use visits of multiple participants within a certain kilometre radius, filtering others out (e.g. Figure 11(A)) and identifying these in the spatial view (Figure 11(B)). This helps identify co-located participants in the same land-use type within certain kilometres (Q4) and adds geographical information to a view otherwise devoid of geography.

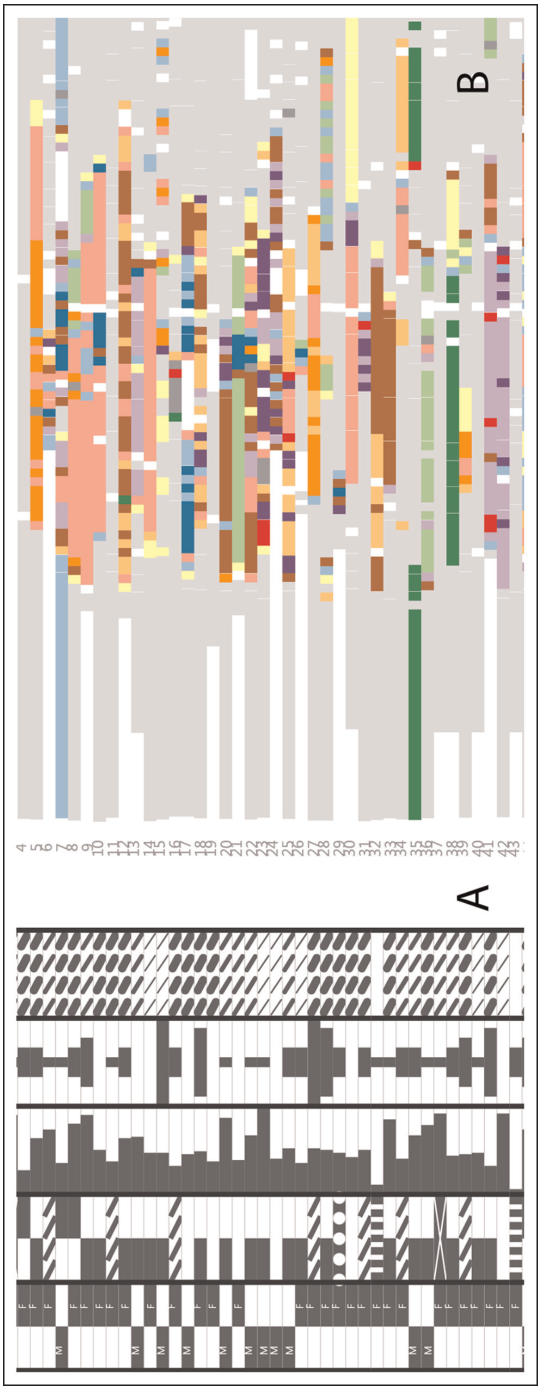

Figure 5. (A) Detail of the participant view, the legend of which is in Figure 6, and (B) zoomed-in portion of the sequence view.

Participant view

The participant view represents socio-economic characteristics of participants to help explain the behaviours observed (Figure 4(B)).[14] It appears to the left of the sequence view, where participants correspond to the appropriate rows of the sequence view and reflect the sorting order. The columns show the five socio-economic characteristics of gender, employment status, ages, income and education from left to right (Figure 5(A)).

Patterns[19] composed of different shapes, sizes, orientations, textures and arrangement are used to denote these socio-economic characteristics (Figure 6). The small cell size makes designing distinguishable patterns difficult, especially by only using variables that are both associative and selective. Therefore, we applied size (only selective) to ordinal values and shapes (only associative) to distinguish nominal categories. We avoided using colour hue or saturation because these overlap with those chosen for the land-use types. We opted for a consistent grey colour to depict these variables instead.

Gender and employment status are nominal. We distinguished gender by filling the left half of the cell for male and filling the right half for females. Distinguishing seven types of employment status encoded by textures, orientations and arrangement (both associative and selective) is difficult. Therefore, seven types of employment status are represented with a pattern based on triangles, dots or lines in different positions and orientations.

Age, income and education are ordinal. Education levels are encoded by textures comprising diagonal lines – the thicker the line, the higher the education level. Age and income are denoted by bar length, left aligned for age and centre aligned for income. Socio-economic details are available on demand when the mouse is moved over a participant.

As stated in the previous section, participants can be sorted on these socio-economic characteristics, enabling us to determine whether participants'

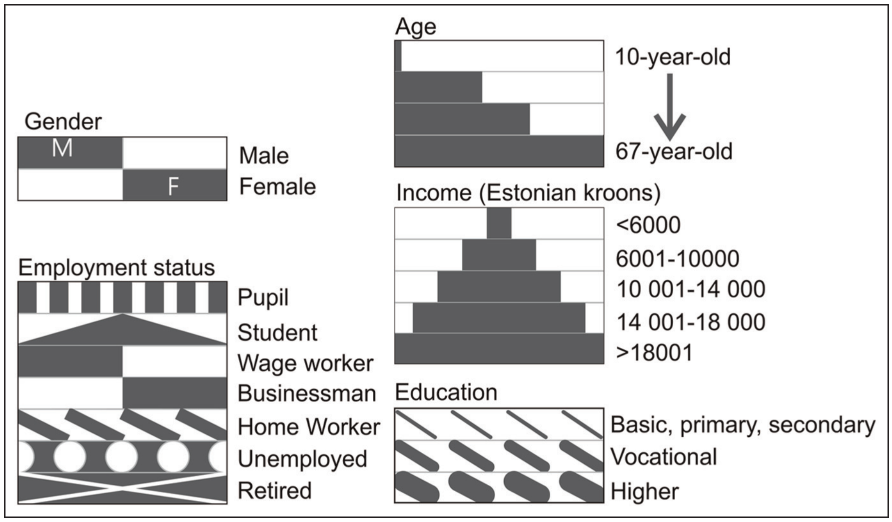

Figure 6. Legend for the participant view.

movements are related to their characteristics (Q7) as shown in Figure 10. Also as stated earlier, dragging the two triangles (to the left of the participant view) along the participant axis allows participants to be filtered (Figure 9(A)).

Connection to the city centre view

The connection to the city centre view ('connection view') highlights when participants enter or leave the city centre (entry and exit events) in three ways: entry/exit times and duration (Figure 4(C-1)), entry/exit counts per day (Figure 4(C-2)) and entry/exit counts per hour (Figure 4(C-3)). This view makes connections visually salient, mainly for identification tasks to support Q1 and Q2. The day can be changed using the tabs positioned on top of this view, comparison among days helps to answer Q6.

Entry/exit times and duration: Participants are arranged from left to right in the same top-bottom order as in the sequence view. The *y*-axis represents 24 h of 1 day. Times in which a participant is in the city centre are shown as grey columns whose bottom and top represent the entry and exit times, respectively. Participants who stayed in the city centre all day long are shown in a lighter grey (Figure 4(C-1)).
Entry/exit counts per day: Accumulated counts for each participant of entry events are shown above the line in dark grey. Counts of exit events are shown below the line in a lighter grey.
Entry/exit counts per hour: This is similar to the daily counts, but entries and exits are aggregated for all participants and are shown per hour. The bar on the left represents accumulated entry events and bar on the right represents accumulated exit events.

Sorting can be applied according to ascending or descending counts of connections to the city centre (e.g. Figure 13(B)). All the views are fully coordinated. Changes in sorting and filtering are reflected in all views.

Spatial view

The zoomable view shows the land-use map (Figure 4(D)). Brushed subsets derived from any other view are superimposed onto the map (e.g. Figure 11), helping depict geographical aspects of the data (Q4).

Dynamic land-use legend

We incorporated participant-level statistics into the legend view (Figure 4(E)).[45] The proportional size (width) of each legend item dynamically depicts the total length of stay over 8 days of each participant identified with the mouse (Figure 7). This view helps localise the land-use types visited by participants (Q5). This view is also linked to other views through brushing.

Implementation

Our design used Excel and ArcGIS 10 for early exploratory ideas and data process. Processing[46] – a programming language conducive to rapid data-driven interactive visualisation design – along with other libraries developed previously[47] was subsequently used to prototype the designs.

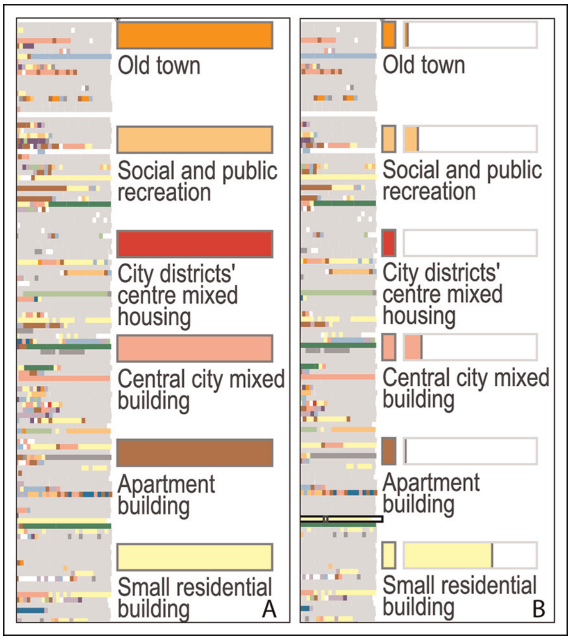

Figure 7. (A) The dynamic legend shows land-use types and (B) the total length of stay over 8 days of the identified participant.

Exploration of user tasks

We now demonstrate the capabilities of the designed visual representations to perform the user tasks. Most of the visual analysis relies on the multiple sorting, filtering, brushing and inspection of all views.

Use of the linked views to answer research questions

Temporal aspects of activities are important for understanding human travel behaviour.[48] The urban study group identified three representative time periods of main activities: 'time for *working*' (10:00–12:00 and 14:00–16:00), '*leisure* time' (19:00–21:00 during weekdays and 12:00–14:00 during weekends) and 'time at *home*' (02:00–04:00). These time periods were used to help identify activities in the visual analysis process.

Identifying patterns through applying multiple sorting in the connection and sequence views is the focus of our first analysis. We first sort participants in ascending order of the number of time they were entering the city on one weekday (day 1). We then sort by land-use types visited by participants at working, leisure and home times on the same day. The result of this four-level multiple sorting is several visually identified clusters of coloured rectangles in the sequence view (Figure 8(B-1) and (C-1)). This indicates that these participants regularly visited the same land uses during weekdays (Q1), showing evidence that regularity exists (Q7). The bar chart at the bottom of the connection view helps us identify how often these connections are made (Q2). Counts to the city are arranged in ascending order from the left to the right in Figure 8(A-2) because frequency patterns from a previous sort were retained.

Q5 requires the complexity of trips to the city centre to be assessed in terms of the consistency of destinations and land-use types. The connection view in Figure 8 shows that the majority of participants have multiple connections to the city during a day. By moving the mouse over participants, the dynamic legend indicates that each participant only visited limited numbers of land-use categories over 8 days. For example, the legend view in Figure 8 shows how long the participant identified with the mouse spent in each land-use type. We conclude that although most participants have multiple connections to the city per day, the land-use types visited are consistent. This explains the temporally repetitive visiting patterns to these land uses.

Q6 requires comparison of temporal changes in connections to the city centre, something facilitated by the connection view. The bar chart to the right of Figure 8(C-2) shows that the majority of participants entered the city between 07:00 and 09:00 and left between 19:00 and 21:00. This can be compared for different days by using the tab to switch between them (Figures 9(B) and 8(C)). Figure 13(B) and (C) shows that participants entered the city later and over a longer time at the weekend than during weekdays (Figure 8), suggesting that participants have a more flexible schedule at weekends.

We visually divided the participants into three groups according to the visual appearances of clusters (colours in the sequence view; Figure 8). Participant group A has fewer visits to the city centre, and these trips are shorter than for other groups. Participant group B shows a greater diversity of visited land-use types during weekdays compared to weekends (the two middle columns). Most participants in group C live near the edge of the city centre, near residential building areas and parks. The land-use types visited during weekdays tend to be diverse than those at weekends.

In Figure 9(A), participants have been filtered such that only those in group B (from Figure 8) remain. Figure 9(B) and (C) shows how these participants are connected to the city centre for the specific days of Thursday and Monday. The similar temporal patterns observed illustrate the temporal periodicity of such connections during working days (Q1).

The second analysis focuses on identifying relations between the patterns and socio-economic

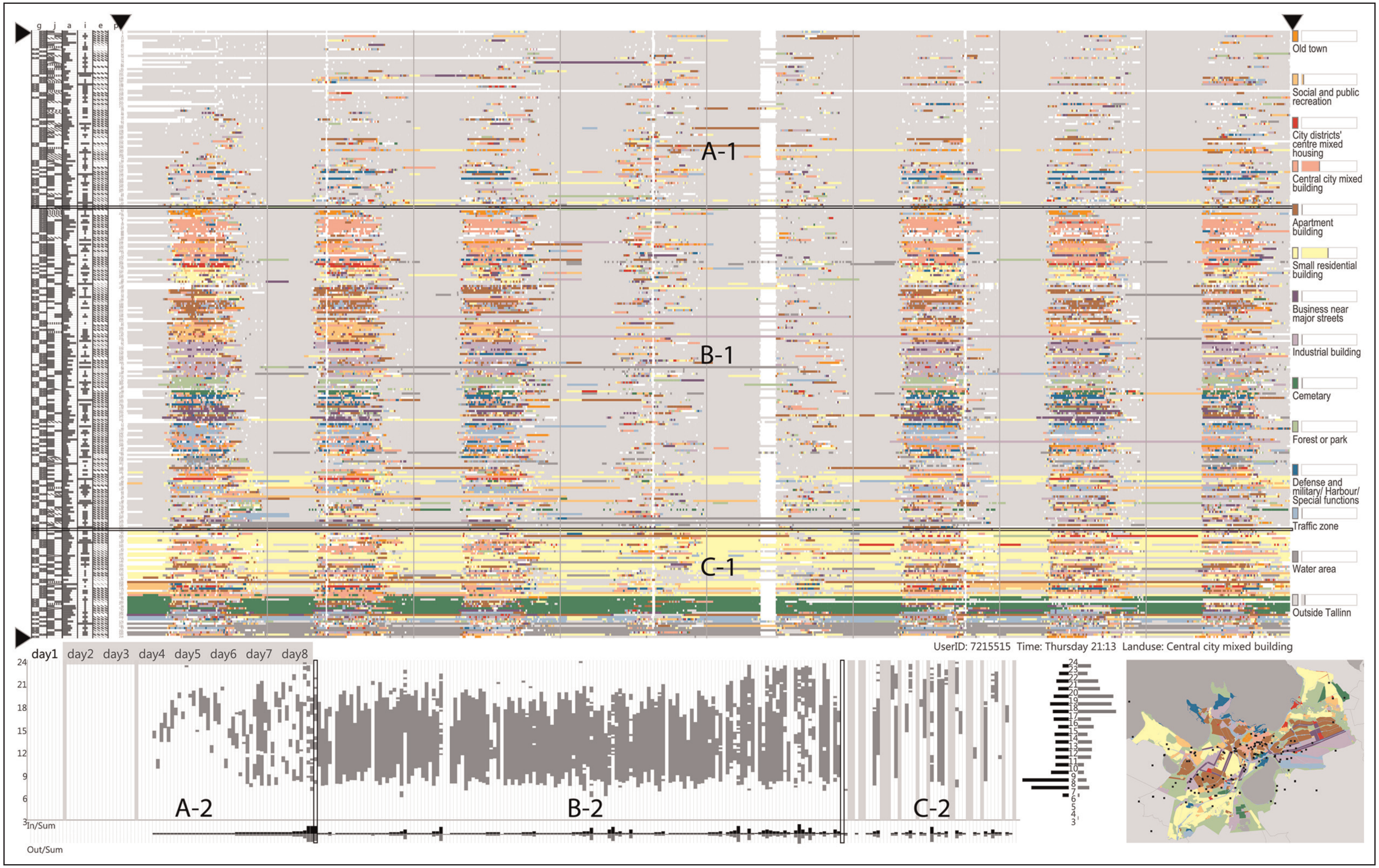

Figure 8. The result of four-level multiple sorting (for day 1) by frequency of city centre entry events, then by land use at 10:00 (work), then by land use at 19:00 (leisure) and finally by land use at 02:00 (home). The results show three (visually determined) clusters of sequences: (A) shows less active patterns of land-use consumption, (B) shows regular patterns during weekdays and (C) shows that participants who live near the edge of the city centre have a greater diversity of land-use visits during weekdays than at weekends. Movement patterns during weekends (two middle columns) can be distinguished from the weekdays in groups B and C.

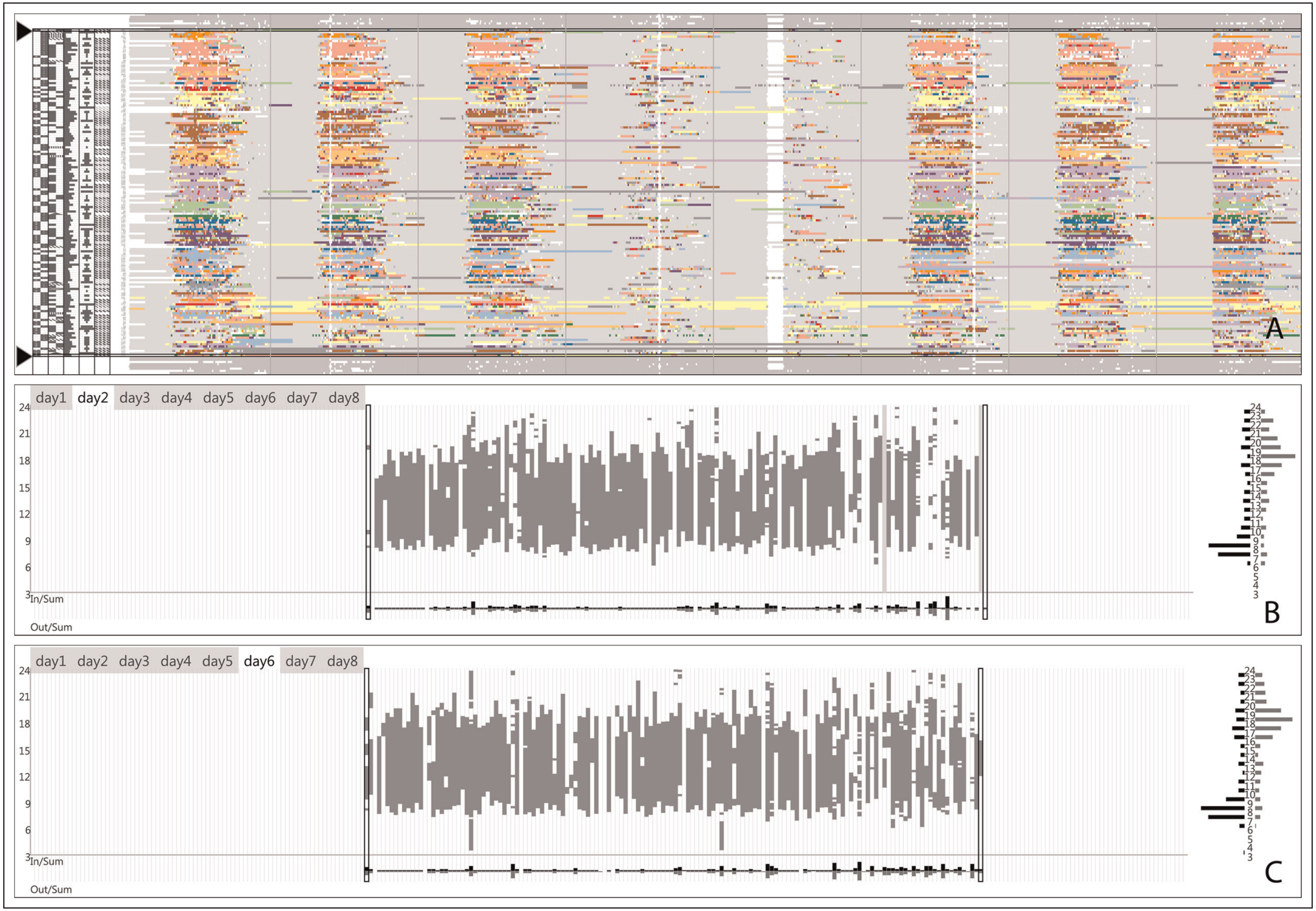

Figure 9. (A) Participants have been filtered by dragging the triangles on the left, so that only those in group B from Figure 8 remain. (B) and (C) Similar patterns of connections to the city centre on Thursday (B) and Monday (C) illustrate weekday periodicity.

characteristics of participants. Q7 asks whether identified patterns can be used to segment participants into different employment groups (second column from the left in the participant view). The three groups of participants we previously identified (A, B and C) do not show direct links with employment categories (Figure 8), although the majority of participants in group B are 'wage workers'.

We further sort the result of Figure 8 by employment type (Figure 10). In this group, 'pupils', 'wage workers' and 'business men' have more regular weekday patterns while the land-use consumption is more homogenous for 'homemakers', 'unemployed' and 'retired' during the whole week.

Q4 asks whether we can localise where people work. We can achieve this for 'wage workers' by finding out where they are during working hours. By selecting the visited land uses by 'wage workers' for working hours and setting the filter parameter to a kilometre (set by the urban study group), we can filter out those who are further than a kilometre away (Figure 11(A)). Using brushing, these are plotted onto the land-use map that can be used to address this research question (Figure 11(B)).

Taking only group 'wage worker' from Figure 10, we can further compare the difference between males and females by sorting them by gender (first column in the participant view). The upper part in Figure 12(A) corresponds to male workers. Female workers spend more time in 'mixed building' areas (pink) during working hours. They have fewer connections to the city centre than their male counterparts, illustrated in Figure 12(B) using Wednesday as an example.

Our third analysis focuses on movement patterns during weekends. We sort participants by land-use types visited at leisure times and home times on the Saturday, and then sort by the number of entry and exit events to the city centre on the same day (Figure 13). As expected, use of land-use types in weekends is more homogenous than during the week, with fewer visits and diverse land-use destinations to the city centre (Q3; Figure 13(A)). The connection view shows that many participants went to the city centre but for shorter time periods. There are no obvious temporal patterns on Saturday (Figure 13(B)) and Sunday (Figure 13(C)). This illustrates that, as expected, trips and total time spent are more flexible at weekends, both in terms of time and space.

Feedback from the urban study group

We consulted the urban study group during development. After completion, we used the prototype with them to address their research questions. An initial chauffeuring approach helped ensure they had a clear understanding of how to use the visualisation through demonstrating techniques and presenting straightforward operational steps.[37] This accelerated their familiarity with the user interface. They then used and tested the prototype with their research questions and the further questions the exploration inevitably prompted.

They reported advantages in using this visual approach for carrying out their *identification tasks*. They enjoyed finding out that the majority of new suburban residents had daily connections to the city centre with more than one trip per day (Figure 8). The prototype helped them interactively deal with a large number of records simultaneously, removing the need to work with smaller subsets. They liked the flexibility to query data for any time interval, instead of being restricted to a predetermined set. They were able to find spatial–temporal patterns identified in the population geography literature.[49–51]

They also found that the prototype was able to support *localisation tasks* relating to the geographical distribution of activities. They found it is easy to localise land-use patterns and their relations with different socio-economic groupings (e.g. Figure 10) and found the visual approach to doing this interesting.

They considered that these approaches to identification and localisation tasks would be applied directly to other aspects of urban research and planning applications, helping them discover connections between different spatial units, finding spatio-temporal patterns related to subgroups and learning about urban structures. They found that the static graphics produced by the prototype were good ways of demonstrating the dynamics of urban life.

After the exploration and tasks were completed by the urban study group, they drew attention to the ability to *compare* visually identify patterns (e.g. Figure 8) based on different characteristics of participants, such as the gender differences showed more female employment in the city centre and a higher female education level attainment (Figure 12). They also drew attention to the ability to *compare* the complexity of trips and found the prototype to have a user-friendly design. They did, however, express the desire to combine these individual-based visualisation methods with more traditional analysis methods.[52] This was particularly true for Q7, for which they suggested that incorporating statistical methods might help them understand how significant differences are.

Discussion

Previous studies have analysed suburban residents' temporal rhythms (by average distances from the city centre and the cumulative sums of daily movement)

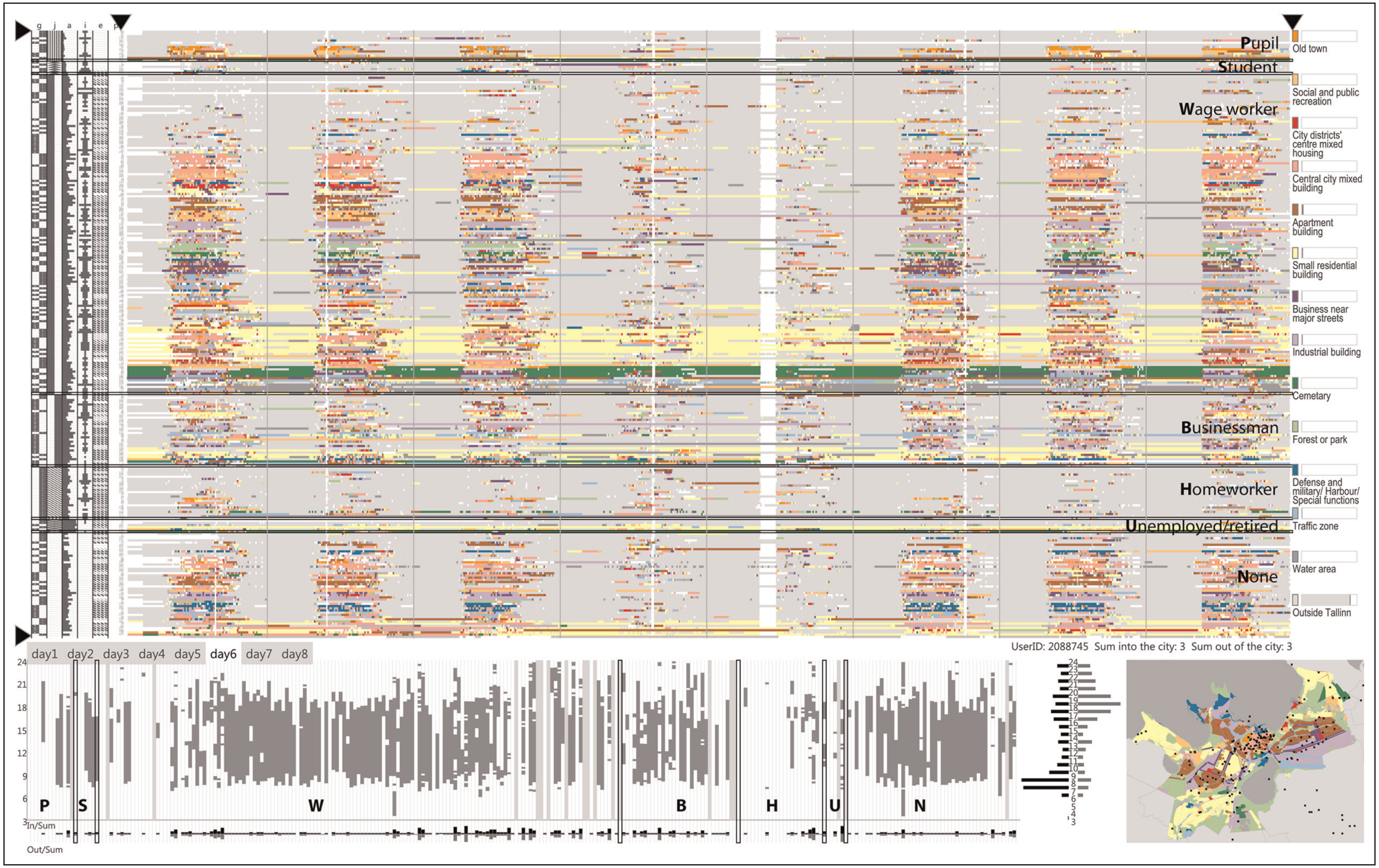

Figure 10. Sorting is applied to the results of Figure 8 by employment status (second column to the left in the participant view) to further group participants into seven groups.

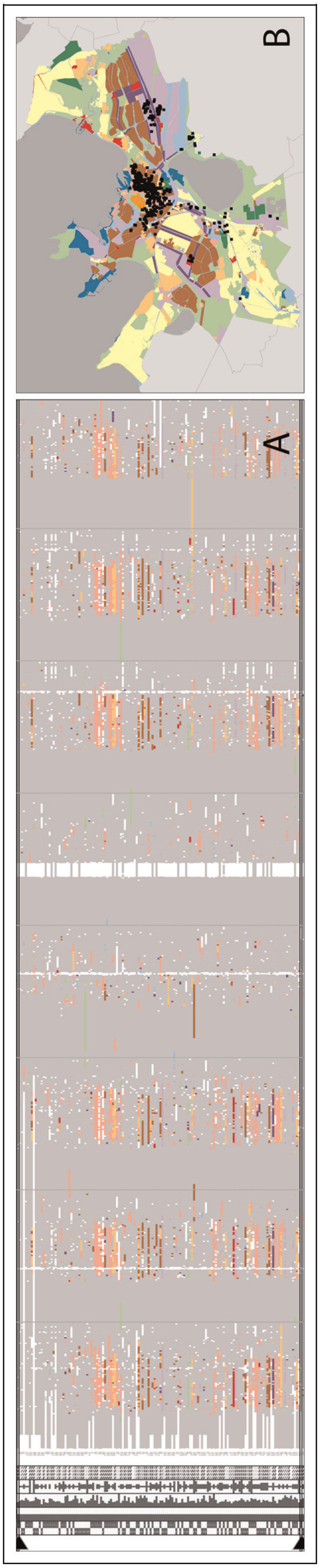

Figure 11. (A) Visited land-use types during working hours are selected. (B) Their geographic locations are superimposed onto the land-use map (in black).

and spatial differences (through daily distribution of the locations in administrative districts) using mobile phone positioning data.[9] The main anchor points (places where participants spent a minimum of 2 h, including work, home and other) were calculated and used in the analysis. Our work incorporates socio-economic characteristics of suburban residents and land-use data to help answer research questions, taking a more flexible, exploratory and visual approach to previous statistical and geographic information system (GIS) analysis studies.

Instead of statistical aggregation, we represent individuals' temporal connections to the city centre and their use of urban space. Our design allows researchers to work on all the data at the same time with filtering where they wish to focus on subsets of the data. The sequence representations allow the land-use contexts (which could be replaced with other contextual data) to be incorporated. Interactions allow the simultaneous analysis on space, time, land use and social aspects, increasing flexibility. Sorting in this case is particularly helpful because it allows the same attributes/values to be gathered together, allowing users to identify and localise patterns.

Our experience of working with the urban study group suggests that a knowledge gap between domain users and visualisation designers exists and that working closely with domain users using iterative design processes can help narrow the gap.[30] Using users' familiar visual representations to explore their data in initial stages helps to accelerate mutual understanding.

Our interactive visualisation explores movements from 277 participants for a week. With this relatively small sample of individuals, it was appropriate to visualise individuals to identify patterns. However, our individual-based visualisation is not scalable to many more participants, because the dimensions of the visualisation cannot be extended and more data may increase cognitive load. A larger sample in large time series will require a more scalable visualisation method or support from computational methods. A larger datasets may need to be aggregated and summarised prior to visualisation, supported with suitable layouts[25] and interactive functions. More sophisticated computational techniques, such as a data mining approach to extract patterns from data, may apply along with the visualisation to deal with large datasets.[6]

We preliminarily tested the design with the urban study group; the observations were confirmed by the urban study group but the conclusions are only indicative as no further formal testing has been conducted. We believe that our design can be applied to other moving objects in other contexts, with a careful consideration of the size of sample and the length of the time period.

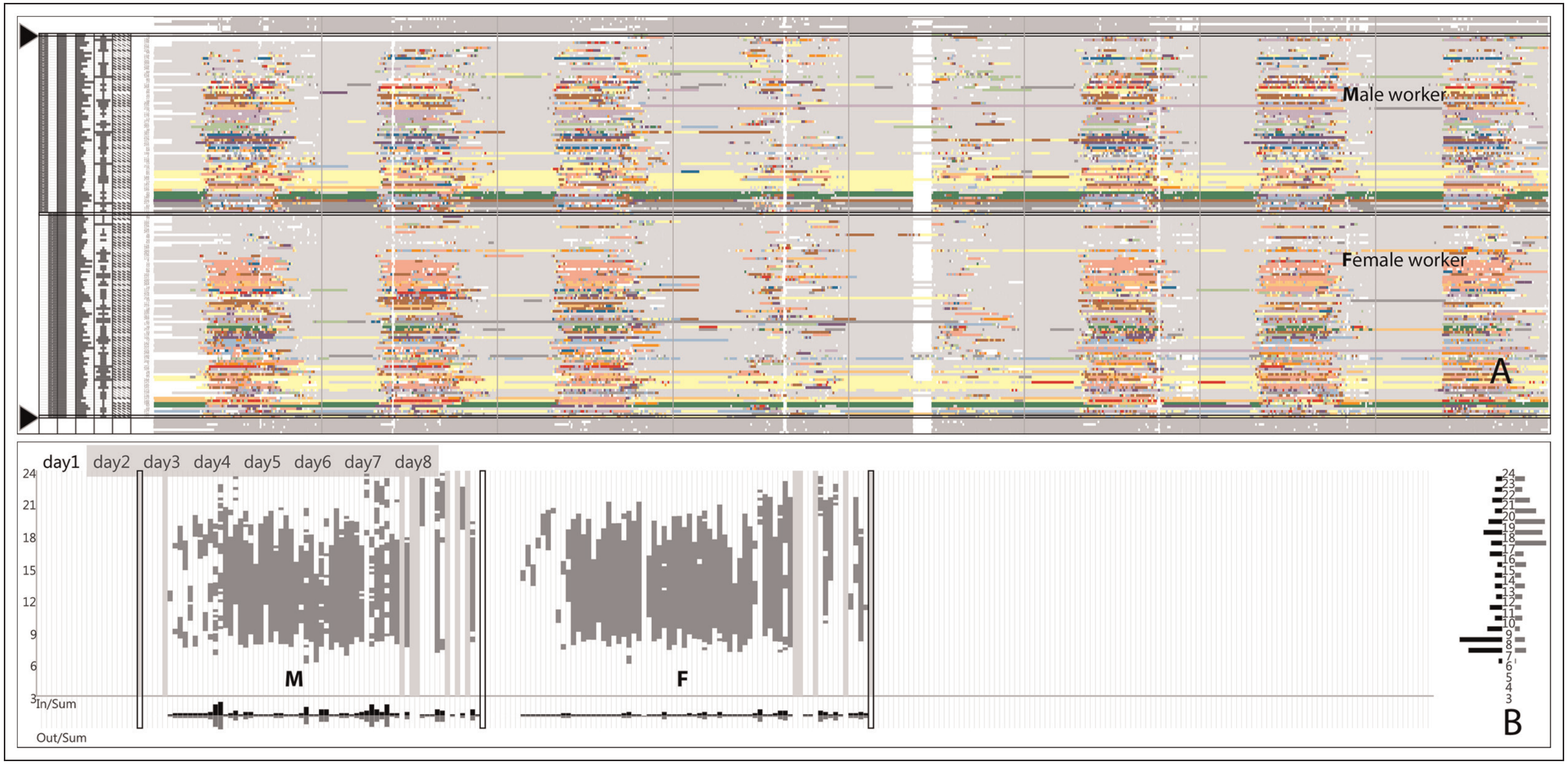

Figure 12. Starting with the result in Figure 10, this groups 'wage worker' by gender. The difference between male worker and female worker is shown in the sequence view (A) and in the connection view (B).

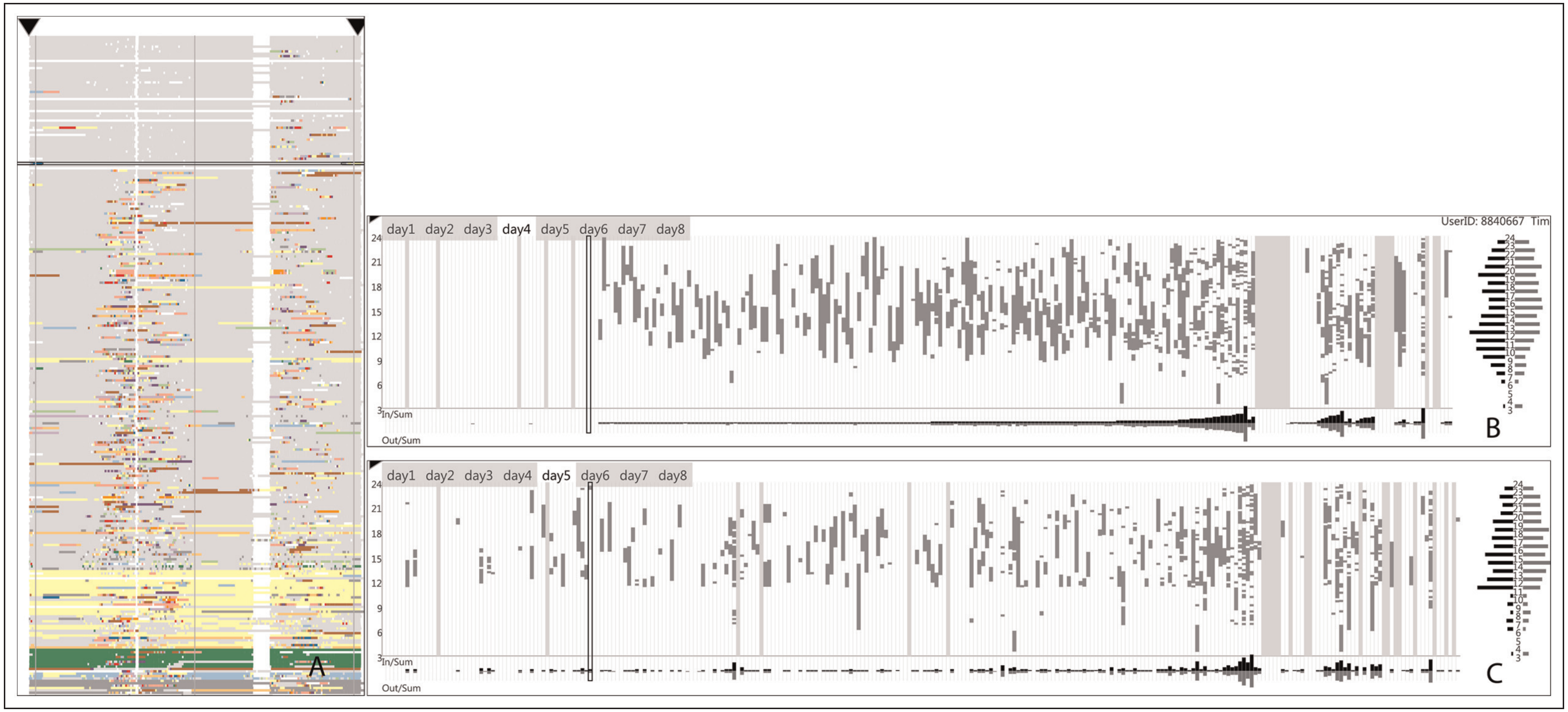

Figure 13. (A) Land-use visits at weekends. Participants are sorted by land-use types by entries/exits to the city centre on Saturday (day 4) and then at leisure times. (B) and (C) Connections to the city centre on Saturday and Sunday, respectively.

Conclusion

This research was motivated by the opportunity to support urban planners in learning about the spatial and temporal dynamics of suburban Tallinn, its connections to the city centre and land-use consumption by residents. The aim of the work was to design interactive visualisation techniques appropriate for exploring and addressing these questions. The design was informed by the research questions of our urban planner collaborators, which were framed as identification, localisation and comparison tasks.

The resulting interactive visualisation tool offers five coordinated and linked views, through which we can identify, localise and compare land-use visiting patterns (sequence view), explore patterns by socio-economic characteristic (participant view), identify temporal patterns of connections between suburbs and the city centre (connection view), superimpose the geographical distribution (spatial view) and show durations (legend view). The urban study group used the resulting prototype to explore their data in the light of their research questions. The coordinated multiple ordering mechanism was found successful to facilitate the visual identification, localisation and comparison of patterns formed from groups of participants with common spatial, temporal and socio-economic characteristics.

We have shown examples of how these designs and techniques enable the research questions to be answered. The results show that the spatio-temporal regularity of movements of suburban residents is associated with their employment status. The urban study group confirmed that the design helped address their research questions and suggested that there was good potential for this approach to support other types of research in their domain. We believe that the taxonomy of tasks that we used and aspects of our design could be usefully applied to research involving movement data in other application domains. Further work is needed to design methods that are more scalable for larger movement datasets in multiple aspects of geo-context.

Funding

We gratefully acknowledge support of the “MOVE Action” funded by the European Cooperation in Science and Technology Programme for facilitating the collaboration and funding Qiuju’s Short Term Scientific Mission that resulted in this work. We also thank the Willis Research Network for funding Aidan’s position.

References

1. Kährik A and Tammaru T. Population composition in new suburban settlements of the Tallinn Metropolitan Area. *Urban Stud* 2008; 45(5–6): 1055–1078.
2. Tammaru T, Leetmaa K, Silm S, et al. Temporal and spatial dynamics of the new residential areas around Tallinn. *Eur Plan Stud* 2009; 17(3): 423–439.
3. Mäe R, Antov D and Antso I. Urban sprawl: mobility potentials in suburban areas of Tallinn. In: Pacetti M, Passerini G, Brebbia CA, et al. (eds) *The sustainable city VII: urban regeneration and sustainability.* Southampton: WIT Press, 2012, pp. 967–977.
4. Ahas R, Aasa A, Silm S, et al. Mobile positioning in space-time behaviour studies: social positioning method experiments in Estonia. *Cartogr Geogr Inform* 2007; 34(4): 259–273.
5. Andrienko G, Andrienko N and Heurich M. An event-based conceptual model for context-aware movement analysis. *Int J Geogr Inf Sci* 2011; 25(9): 1347–1370.
6. Andrienko G, Andrienko N, Dykes J, et al. Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. *Inform Vis* 2008; 7(3–4): 173–180.
7. Keim D, Kohlhammer J, Ellis G, et al. *Mastering the information age: solving problems with visual analytics*. Goslar: Eurographics Association, 2010, p. 168.
8. Tory M and Moller T. Human factors in visualization research. *IEEE T Vis Comput Gr* 2004; 10(1): 72–84.
9. Ahas R, Aasa A, Silm S, et al. Daily rhythms of suburban commuters’ movements in the Tallinn Metropolitan Area: case study with mobile positioning data. *Transp Res Part C Emerg Technol* 2010; 18(1): 45–54.
10. Tammaru T. Suburbanisation, employment change, and commuting in the Tallinn Metropolitan Area. *Environ Plann A* 2005; 37(9): 1669–1687.
11. Ahas R, Laineste J, Aasa A, et al. The spatial accuracy of mobile positioning: some experiences with geographical studies in Estonia. In: Gartner G, Cartwright W and Peterson MP (eds) *Location based services and telecartography.* Berlin, Heidelberg: Springer, 2007, pp. 445–460.
12. Kang C, Gao S, Lin X, et al. Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In: *18th international conference on geoinformatics*, Beijing, China, 18–20 June 2010, pp. 1–7. New York: IEEE.
13. Andrienko G, Andrienko N, Bak P, et al. A framework for using self-organising maps to analyse spatio-temporal patterns, exemplified by analysis of mobile phone usage. *J Locat Based Serv* 2010; 4(3): 200–221.
14. Slingsby A, Beecham R and Wood J. Visual analysis of social networks in space and time. In: *The Nokia data challenge workshop (Pervasive 2012)*, Newcastle, 18–19 June 2012.
15. Eagle N and Pentland A. Reality mining: sensing complex social systems. *Pers Ubiquit Comput* 2006; 10(4): 255–268.
16. Dal Fiore F, Beinat E, Rietveld P, et al. Exploring the relationship between land use and mobile phone usage: analysis of a dataset for the Amsterdam metropolitan region. In: *Workshop on pervasive urban applications (PURBA)*, San Francisco, CA, 12–15 June 2011.
17. Skupin A and Miller HJ. Nokia MDC atlas: an exploration of mobile phone users, land cover, time, and space.

In: *Nokia data challenge workshop (Pervasive 2012)*, Newcastle, 18–19 June 2012.
18. Di Lorenzo G and Calabrese F. Identifying human spatio-temporal activity patterns from mobile-phone traces. In: *14th international IEEE conference on intelligent transportation systems (ITSC)*, The George Washington Univeristy, Washington, DC, USA, 5–7 October 2011, pp. 1069–1074. Washington: Omnipress.
19. MacEachren AM. *How maps work: representation, visualization and design*. New York: The Guilford Press, 1995, p. 513.
20. Wehrend S and Lewis C. A problem-oriented classification of visualization techniques. In: *Proceedings of the 1st conference on visualization '90*, San Francisco, CA, 23–26 October 1990, pp. 139–143. Los Alamitos, CA: IEEE Computer Society Press.
21. Knapp L. A task analysis approach to the visualization of geographic data. In: Nyerges TL, Mark DM, Laurini R et al. (eds) *Cognitive aspects of human-computer interaction for geographic information systems*. Dordrecht: Kluwer Academic, 1995, pp. 355–371.
22. Bertin J. *Semiology of graphics: diagrams, networks, maps*. Madison, WI: The University of Wisconsin Press, 1983, p. 415.
23. Andrienko N and Andrienko G. *Exploratory analysis of spatial and temporal data: a systematic approach*. Berlin: Springer, 2006, p. 703.
24. Friendly M and Kwan E. Effect ordering for data displays. *Comput Stat Data An* 2003; 43(4): 509–539.
25. Slingsby A, Dykes J and Wood J. Configuring hierarchical layouts to address research questions. *IEEE T Vis Comput Gr* 2009; 15(6): 977–984.
26. Fabrikant SI. Building task-ontologies for geovisualization. In: *ICA commission on visualization and virtual environments pre-conference workshop on geovisualization on the web*. Beijing, China, 3–4 August 2001.
27. Tallinn Department of Planning. Tallinn city masterplan (Tallinna linna üldplaneering), http://www.tallinn.ee/est/ehitus/Tallinna-linna-uldplaneering (accessed 29 December 2012).
28. Brewer CA. Chapter 7: color use guidelines for mapping and visualization. In: MacEachren AM and Taylor MAP (eds) *Visualization in modern cartography*. Tarrytown, NY: Elsevier Science, 1994, pp. 123–147.
29. Koh LC, Slingsby A, Dykes J, et al. Developing and applying a user-centered model for the design and implementation of information visualization tools. In: *15th international conference on information visualisation (IV)*, London, United Kingdom, 13–15 July 2011, pp. 90–95. Los Alamitos, CA: IEEE Computer Society Press.
30. Slingsby A and Dykes J. Experiences in involving analysts in visualisation design. Beyond time and errors: novel evaluation methods for visualization (BELIV). In: *Visweek 2012*, Seattle, WA, 14–15 October 2012.
31. Aigner W, Miksch S, Schumann H, et al. *Visualization of time-oriented data*. London, Dordrecht, Heidelberg, New York: Springer, 2011.
32. Lloyd D and Dykes J. Human-centered approaches in geovisualization design: investigating multiple methods through a long-term case study. *IEEE T Vis Comput Gr* 2011; 17(12): 2498–2507.
33. Roberts JC. State of the art: coordinated & multiple views in exploratory visualization. In: *Fifth international conference on coordinated and multiple views in exploratory visualization (CMV '07)*, Washington, DC, 2 July 2007, pp. 61–71. Los Alamitos, CA: IEEE Computer Society.
34. Robertson G, Fernandez R, Fisher D, et al. Effectiveness of animation in trend visualization. *IEEE T Vis Comput Gr* 2008; 14(6): 1325–1332.
35. Morrison JL. A theoretical framework for cartographic generalization with the emphasis on the process of symbolization. *Int Yearb Cartogr* 1974; 1974(14): 115–127.
36. Ware C. *Information visualization: perception for design*. 2nd ed. Amsterdam: Elsevier Morgan Kaufmann, 2004, p. 486.
37. Slingsby A, Dykes J and Wood J. Exploring uncertainty in geodemographics with interactive graphics. *IEEE T Vis Comput Gr* 2011; 17(12): 2545–2554.
38. Becker RA and Cleveland WS. Brushing scatterplots. *Technometrics* 1987; 29(2): 127–142.
39. Yi JS, Ya Kang, Stasko JT, et al. Toward a deeper understanding of the role of interaction in information visualization. *IEEE T Vis Comput Gr* 2007; 13(6): 1224–1231.
40. Shneiderman B. The eyes have it: a task by data type taxonomy for information visualization. In: *Proceedings of IEEE symposium on visual languages*. Boulder, CO, 3–6 September 1996, pp. 336–343. Los Alamitos, CA: IEEE Computer Society.
41. Shneiderman B. Tree visualization with tree-maps: 2D space-filling approach. *ACM T Graphics* 1992; 11(1): 92–99.
42. Becker RA, Cleveland WS and Shyu M-J. The visual design and control of trellis display. *J Comput Graph Stat* 1996; 5(2): 123–155.
43. Friendly M. Mosaic displays for multi-way contingency tables. *J Am Stat Assoc* 1994; 89(425): 190–200.
44. Gleicher M, Albers D, Walker R, et al. Visual comparison for information visualization. *Infor Vis* 2011; 10(4): 289–309.
45. Dykes J, Wood J and Slingsby A. Rethinking map legends with visualization. *IEEE T Vis Comput Gr* 2010; 16(6): 890–899.
46. Fry B. Processing, http://processing.org/ (accessed 29 August 2012).
47. Wood J, Slingsby A and Dykes J. giCentreUtils, http://gicentre.org/utils/ (accessed 29 August 2012).
48. Harvey AS. From activities to activity settings: behaviour in context. In: Ettema D and Timmermans H (eds) *Activity-based approaches to travel analysis*. Tarrytown, NY: Elsevier Science, 1997, pp. 209–228.
49. Hugo G. Population geography. *Prog Hum Geog* 2007; 31(1): 77–88.
50. Kwan M-P. Gender differences in space-time constraints. *Area* 2000; 32(2): 145–156.
51. Bromley RDF, Tallon AR and Thomas CJ. Disaggregating the space-time layers of city-centre activities and their users. *Environ Plann A* 2003; 35(10): 1831–1851.
52. Ahas R and Mark Ü. Location based services–new challenges for planning and public administration? *Futures* 2005; 37(6): 547–561.